Biases in categorization

Das-Smaal, E.A.

published in
Cognitive Biases
1990

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Das-Smaal, E. A. (1990). Biases in categorization. In *Cognitive Biases* (pp. 349-386). Elsevier Science Publishers B.V..

Cognitive Biases
J.-P. Caverni, J.-M. Fabre and M. Gonzalez (Eds.)
© Elsevier Science Publishers B.V. (North-Holland), 1990

BIASES IN CATEGORIZATION

Edith A. Das-Smaal

Department of Cognitive Psychology, Vrije University
Amsterdam, The Netherlands

Abstract. On what grounds can we conclude that an act of categorization is biased? In this chapter, it is contended that in the absence of objective norms of what categories actually are, biases in categorization can only be specified in relation to theoretical understandings of categorization. Therefore, the chapter starts with an overview of current theoretical positions regarding categorization, and the various constraints on learning and using categories that are specified by these theories. Several types of categorization bias follow from these descriptions. In this context, the role of prototypes, basic level categories, existing schemata, category width, context, feature frequency and correlation among features is discussed, as well as the biasing influence of selective attention, hypothesis testing and task demands.

1. INTRODUCTION

1.1 On categories and reality

This discussion of biases in categorization will be preceded by a consideration of the question of whether biases in this area can be established at all. When can we state that an act of categorization is biased? Biases are defined as systematic error, or the difference between judgments and the true value (e.g. Anderson *et al.*, 1981). In order to be able to claim that an observed categorization behavior is biased, it is therefore necessary to specify what categories, in reality, are like. However, features, objects and events of the real world can be categorized in an endless number of ways. Besides, our perception is highly selective and thus biased already. Which features we do perceive will be influenced by factors like for instance our physical ability to perceive them, and by our needs, our motives and prior knowledge, as represented in our cognitive system.

It is at present widely assumed that categories arise from interaction between events in the world and the perceiver with his own prior knowledge. The perceiver, in making sense out of the stimuli of the world, has his or her own contribution to the resulting representation of the world. As McCauley (1987) states, from their mental constructs or models, people often impose categories on the world of their experience rather than induce them from this world, although some categories seem to have an ecological basis too. It is important to recognize that human categorization behavior is heavily dependent on the cognitive constructs, the models that are built up during a lifetime of experience. Specifying the real world as it is, not as we perceive and represent it, is simply beyond our reach.

So in the case of categorization, the norm has to be a question of theory, because there is no way of being sure that the real world is structured as we perceive it, with categories existing *a priori*. Perception of the world is constrained by our mental constructs, which in turn influence our categorization performance. However, there are a variety of theories and models on categorization, and although there seems to be some consensus about which theories have to be abandoned by now, within these constraints there is much less consensus about the appropriate description of human categorization behavior. An example on the issue of prototype effects may serve to illustrate this. As we shall see in the next paragraph, from a classical point of view, all members of a category should be as representative of the category as any other. This follows from the assumption that the norm for category membership is sufficiently specified just by the presence or absence of a set of defining features. All other features, even if they are representative of the category, are considered irrelevant, and this makes all category members equivalent. However, advances in our understanding of categorization have made it clear that in many categories differences in typicality do exist among category exemplars. People do not judge all category members as equally representative of their category. Thus, systematic departures appear to occur from the definition of what constitutes a category in a classical point of view. However, in a prototype conception typicality effects are not biases in this sense. In this view they are in the core assumption of the theory, as will be outlined below. Prototype views are specially devised to account for typicality effects, trying to capture why it is that membership is a matter of degree. So, in the process of theoretical development, categorization behavior that appeared to be biased may prove to be unbiased given another theoretical understanding. Theoretical development keeps going on. Again, the issue of typicality effects may illustrate this. Typicality effects, in a still more recent theoretical account, are considered as phenomena that are not necessarily relevant to categorization. It is contended that they need not show something direct about the nature of categorization. They are superficial effects that do not always mirror category structure directly. Prototype effects may result from many sources, and are to be accounted for in terms of deviations from one's cognitive models or "theories" for understanding the world (Lakoff, 1986). We shall return to this view later.

The point here is that misinterpretations, or at least different kinds of interpretations of categorization behavior, continue to occur, due to an incomplete understanding of all the types of categories that occur, and to a lack of knowledge regarding all aspects of influence on categorization. One of the important aspects is people's prior knowledge. Another point to mention is that in categorization models, tasks and situations have not received much attention. They have nearly always implicitly been considered to be homogeneous. That is, models usually do not predict different categorization behavior in different tasks and contexts. It is true, however, that tasks as well as situations do influence categorization, as will be shown in this chapter.

In sum, absolute or clearly established standards, from which departures can be described, unfortunately do not seem to be at hand. However, there is a sense in which biases in categorization can be specified. It is obvious that there are a variety of potential groupings of the things in the world around us. Likewise, there are many potential ways of representing such groupings, and ways of deciding on group membership. An important question, therefore, is which ones are actually done or preferred by us and which ones are not. Preferential biases can be studied experimentally by investigating constraints in human categorization activities. To this goal, designing experimental categories so as to exclude ambiguous interpretations of categorization behavior can be profitable. Preferred ways of categorizing will provide important clues as to both the underlying structure and the process of categorization. For instance, people may show a "bias" favoring correlated features, as was suggested by Rosch (1978). Directing attention towards certain features and away from others can be another source of bias, pertaining to what features are represented when a category is learned. Or people may or may not consider the occurrence of features in contrast categories when making a categorization decision. If biases in categorization are conceived of as preferential ways of learning and using categories, both these ways and the factors of influence on them have to be described. As pointed out in this opening paragraph, it should be kept in mind that this conception of biases ultimately boils down to presenting a descriptive theory of human categorization behavior. Despite that, the approach that is taken in this chapter is not to present a full theory of categorization, but rather to describe biases as constraints with which models should not be inconsistent.

Bearing this understanding of biases in mind, we shall deal with the subject of categorization in general first, and then turn to a discussion of a number of important biases in this matter. It is our theoretical point of view that biases can result from limitations on aspects of category information that are stored in memory, or from computational constraints whenever that categorical knowledge is used. Both aspects will be considered in this chapter.

1.2 Categorization

One of the most fundamental aspects of thinking is the ability to perceive similarities and differences among the events that we experience. Events rarely repeat themselves in exactly the same way. And even if they did, it is still questionable whether our interpretation would be quite the same as before. In this respect, impressions are always new to us. It is our conceptual system, however, that allows us to perceive similarities between new and old experiences. We are inclined to relate incoming information to things that we know. One mental operation by which this can be accomplished is categorization. By categorizing, the continuous variation in environmental information is reduced to manageable and knowable proportions. Categorization provides a means of attaining cognitive economy (Rosch, 1978). In this way, adequate reactions are possible.

Mental categories generally are formed by experience. Experience, however, usually encompasses only a reduced sample of the total set of exemplars of a category, and this raises a number of important questions. The main issue, of course, is the question of what these mental categories are like. What is the nature of the process of categorization? Are there general constraints regarding this process? How can limited experience lead to knowledge of the whole category? What are the constraints on this knowledge? And how, for instance, do we determine the boundaries of a category? How is experience related to category width? A variety of answers have been given to these and related questions on how we form a mental representation of a category on the basis of our experience. A short description of the historical development of views on categorization and the main current theoretical positions will be dealt with first. From that, it follows that categorization is biased in a number of different ways. We shall go into this matter subsequently. With respect to biases, in succession, the role of prototypes, basic level categories, existing schemata, the formation of category boundaries and the influence of context, the representation of feature frequency and correlation among features will be discussed, as well as the issues of selective attention, hypothesis testing and task demands.

2. THEORETICAL VIEWS ON CATEGORIZATION

2.1 Historical development

The study of categorization emerged within the field of psychological learning theory. A broad distinction can be made between earlier "associationistic" stimulus-response (S-R) theories and later hypothesis-testing theories. It was the post-1950's tradition of cognitive psychology that led to a preference for active hypothesis-testing models. The behaviorist tradition prior to the cognitive shift entailed a passive associationistic account of the category learning process. Because of the hypothesis-testing theories the historical development will be briefly described here. Hypothesis testing directs attention to certain aspects of the information encountered, and this can be assumed to have a biasing influence on the acquisition of categories. Besides, in hypothesis testing people may adopt various strategies, and show biases in this respect too.

As for the S-R account, it was Hull (1920) who advanced the "common element" view of category concepts. In his theory, category learning involves the formation of an association between a certain response and an element common to a class of different stimuli. The category learner is viewed as passively undergoing a series of experiences which gradually bring about these associations, by reinforcement. Hull's opinions were formed in the behavioristic climate of his time. In this tradition the unobservable internal processes of the organism were ignored and considered as a subject unsuitable for scientific enterprise. Throughout the 1950's, the passive and mechanical associationistic account of the category learning process prevailed. S-R theories, however, appeared to be too simplistic to account for central or thought processes that may influence the category learner's response to the information presented in the stimulus.

In about 1955, the S-R concept was to replaced by the input-output concept. It was recognized that a complex program controlled the input-output sequence (Miller, Galanter, & Pribram, 1960), and this opened the way for new approaches to the study of cognitive processes, leading to the discovery of some sources of bias in categorization. A theory was proposed according to which the organism is seen as actively seeking information. According to this view, categories are learned via an active, strategic process of forming and testing hypotheses. Bruner, Goodnow and Austin (1956) advanced this view. They provided an analysis of the nature of categorization and its central role in cognition. They tried to externalize the thought processes of people when learning to categorize. They broke with the behavioristic tradition by verifying the existence of hypothesis testing behavior. From the observed behavior of the subjects they inferred certain patterns of purposive behavior which they called strategies. The notion of strategies enabled them to describe what the subject does "internally" when learning new classifications.

The work of Bruner *et al.* brought about fundamental changes in category learning theory, and stimulated an intensive analysis of the variables affecting and the processes underlying category learning. The process by which subjects learn a category was accepted to be one of forming hypotheses, which are tested and revised in the light of information from experience. It was realized that categories generally are learned by examples. People often have no control over which instances are presented to them. Exemplars and non-exemplars are pointed out to them as they happen to occur. In that case, a logical way to learn the category is by following a kind of reception strategy. This involves adopting a hypothesis, based on either the whole or a part of the first category example, and knowing how and when to alter it whenever new information is presented. To study the strategies actually adopted by subjects, Bruner designed a task that has since been used as a standard procedure. Visual stimuli were constructed that subjects had to learn to classify into categories. Categories were defined by the experimenter. For example, the subject was shown a series of geometric figures that differed along the following dimensions: form (the values being circle and square), colour (red and blue) and size (small and large). A category could then be defined by all red circles, for instance, leaving size as an irrelevant dimension. The subject was asked to categorize each figure presented. In a reception paradigm, the experimenter determined which stimulus was shown, whereas in a selection paradigm the subjects were to choose the next figure themselves. Following each categorization the subject was told whether or not the response was correct. Learning took place from this feedback information. It was inferred that the subject had identified the category when no more categorization errors were made. Learning trials continued until the subject consistently responded correctly.

Bruner *et al.* showed that in this task there is a bias regarding strategies of category learning. People do not behave randomly, but appear to prefer certain actions above others in a systematical way. Most people choose one of two strategies in formulating hypotheses and selecting subsequent stimuli for testing as far as a selection paradigm is concerned. A sizable number of subjects adopt a wholistic or focusing strategy. That is, they formulate a global, composite hypothesis based on the first category member they encounter, and use this hypothesized rule in making subsequent responses until they encounter an instance that refutes the rule. Another strategy, adopted less frequently, is to formulate hypotheses involving only a limited number of features of the first example presented, rather than all of them. Such a strategy is called scanning. The former strategy seems to be the optimal one in this type of task, because the demands it makes on the subject's memory are not as high as the latter. In general, people show a preference for the strategy that requires the least memory load, that is, focusing. Nonetheless, the scanning strategy was also found to be used in Bruner's experiments. With the paradigm mentioned above, other issues were investigated too. The influence of stimulus factors like the number of relevant and/or irrelevant dimensions, the number of values on the dimensions, and the type and complexity of the rule defining the category, was studied, as were procedural factors such as the order of presentation of the stimuli and the amount and nature of feedback to the subject. The investigations also dealt with the way in which hypotheses are revised on the basis of feedback.

Many of these variables and their combinations have been shown to affect categorization learning (Bourne, Dominowski, & Loftus, 1979). Categories defined by different rules relating the relevant features appeared to vary in difficulty. Also, it generally appeared that subjects enter a category learning problem with a bias favouring conjunctive (e.g. red *and* circle) over disjunctive (red *and/or* circle) rules. Subjects more frequently state conjunctive rather than disjunctive categorization rules when both are possible (Hunt & Hovland, 1960), and irrespective whether a conjunctive or a disjunctive rule applies, they tend to start testing conjunctive ones (De Swart & Das-Smaal, 1976). However, as we shall see with many factors in categorization, rule difficulty is not quite independent of the particular stimulus context in which it occurs. Pairs of features may or may not go together naturally, and this influences the ease with which either conjunctions or disjunctions are formed (Dominowski & Wetherick, 1976: Reznick & Richman, 1976). Regarding the search for category rules, Martin and Caramazza (1980) in a more recent experiment presented evidence that subjects develop rule systems also when learning less well-defined categories. In their experiment, there was no simple set of features that could be used to determine membership of all exemplars of the category. In that case, subjects appeared to look for rules that would enable them to categorize the stimuli as easily as possible. Because the categories were structured so that no simple rule could serve to categorize the stimuli, subjects were forced to develop complex sets of rules if indeed they preferred to develop rules. Martin and Caramazza showed that this was what they actually did, even in a complex situation. We shall return to rule development and hypothesis testing as a source of bias later, in connection with selective attention.

The studies mentioned above showed the existence of bias regarding strategies of category learning and the search for categorization rules. The question of the nature of the information that is stored in memory when a category is learned did not receive much attention in the earlier studies. It was taken for granted that categories simply consist of "defining" values and that categorization is based on rules for combining these values. The emphasis was placed on the learning process rather than on structure. In about the last fifteen years, the trend of research on categorization changed towards focusing on the issue of what is stored in memory, thereby questioning the idea of defining value representation. The shift went together with more interest in natural, real-life categories.

2.2 Conceptions of categories

According to what was to be called the "classical" conception of categories mentioned above, it was assumed that all members of a category share one or more defining features. If there is more than one common feature, then a certain relationship among the features is specified in the definition. Thus, in this view, each category has its own definition that provides necessary and sufficient conditions for assigning category membership. Features that are not included in the definition are irrelevant to category membership. By experience we learn which features are the ones that define the category, and by what rule they are connected. Following learning, the categorization decision about new items is based exclusively on whether or not an item fulfills the conditions in the definition, and this is an all-or-none matter. This way, the category boundaries are determinate and well-defined. Also in this view, any one exemplar of the category should be as representative of the category as any other. An example of a well-defined category is the category of triangles, or sisters-in-law, or odd numbers.

However, not all categories are that strictly definable. Many categories may not conform to the classical view. It appears, for instance, that in all kind of tasks non-defining features of a category can also play an important role (e.g. Rosch, 1975; Das-Smaal & De Swart, 1984, 1986). Non-defining features may contribute to differences in typicality among category exemplars. An example is the category of fruit. An apple is, at least to a Westerner, a better example of the category of fruit than a mango. Furthermore, the category boundaries are not always quite clear. Not everyone is certain about whether a tomato, a coconut or an olive should be counted as a fruit. The boundaries between categories like fruit and vegetables are vague. Vague boundaries and typicality differences are not in keeping with the classical view of categories. This has led many investigators to alternative views, mainly prototype ones. In the prototype approach, the representation of a category is also taken to be the result of an abstraction process, and it is generally assumed that different category exemplars are integrated into one memory structure, i.e. a summary representation of the category. However, in contrast to the traditional view, it is thought that category representation is not restricted to a set of defining features. For a feature to be included in the summary, it need only be characteristic of the category.

It was Rosch's work (1973, 1975) that applied prototype theory to natural categories, the implicit concepts of daily life. Rosch and her co-workers have demonstrated the existence of prototype effects in these categories. They showed that natural categories have a graded structure in that some members are usually regarded as prototypes or typical exemplars, whereas others are considered less typical of the category. More specifically, Rosch hypothesized that typicality of a category member is determined by its family resemblance to other members of the same category. A high family resemblance means that a large number of features are shared with the other members, while at the same time few features are shared with contrasting categories. Thus, the best exemplars of one category will not be good representatives of other comparable categories.

Typicality effects not only occur in natural categories like fruit and vegetables, but also in ad hoc constructed and goal-derived categories (Barsalou, 1983, 1985). In addition, typicality effects are also reported to occur in well-defined categories. Das-Smaal & De Swart (1981) were able to show this for a number of different tasks. Categories that were constructed according to the traditional view -- with membership being sufficiently specified just by defining features -- were nevertheless learned in a way that favoured typical members over less typical ones. Bourne (1982), Armstrong, Gleitman, & Gleitman (1983) and Vandierendonck (1989) also demonstrated typicality effects in classically defined categories like for instance odd numbers. A view according to which no aspect of within-category variation is represented does not account for these effects. A prototype position offers a better explanation of the results.

There are a number of different interpretations of prototype effects. A fairly common view, following the work of Rosch (1973, 1975), is that a prototype is an abstraction, an integration of exemplars in some kind of memory structure. Category membership can be gradual, and degree of membership of an exemplar is determined by its similarity to the prototype. Later, Rosch (1978) herself became somewhat reserved on the issue of prototypes as a theory. Her position by 1978 was that prototype effects do indicate that prototypes, or best examples, must have some place in psychological theories of representation, processing, and learning, but that they do not in themselves constitute any particular representation or process model. This point is stressed by Lakoff (1986). Lakoff, however, goes further and states that although he acknowledges that prototype effects are real and that even classically defined categories may give rise to prototype effects, such effects are mere shadows of cognitive models. They show nothing direct about the nature of categorization. Prototype effects are a result of the fact that knowledge is organized in terms of cognitive models. Cognitive models are used to structure and make sense out of our experience. There are various kinds of cognitive models, and hence prototype effects may come from different sources. Although direct correlations between conceptual structure and prototype effects do exist, in other cases the effects may arise indirectly. They do not necessarily correspond to degrees of membership of exemplars. A full study of category structure therefore must provide an account of the details of the cognitive models that give rise to the effects.

Other "post-Roschian" theoretical proposals also stress the importance of theories or mental models in categorization. In general, it can be stated that since Rosch's original publications, the theoretical thinking about categorization has been moving away from an emphasis on objective features and similarity to an insistence on the role of cognitive models in categorization behavior. The influence of cognitive models provides an obvious source of bias, which will be discussed later in the sections on prototypes and schemas. Suffice it to say here that theoretical development will undoubtedly lead to a richer view of intra- and inter-concept relationships than that advanced so far (e.g. Medin & Wattenmaker, 1987; Neisser, 1987).

Besides the classical and prototype approaches, a third conception of categories is the exemplar view. According to this idea, categories are represented by the collection of individual exemplars someone knows. In this case there is no abstracted representation of a category. Rather, categorical knowledge consists of an extensional description of the set of experienced exemplars. Categorization of an unknown exemplar is determined by the degree to which it is similar to an already experienced exemplar. The weaknesses of the exemplar view were outlined by Das-Smaal (1986). One argument against exemplar representation is for instance the fact that specific exemplars are often poorly remembered (e.g. Bourne & O'Banion, 1969). The exemplar view also encounters problems in representing generalized knowledge that pertains to exemplar features, such as knowledge on conjoint frequencies or relevancy of features to the categorization decision. As for the latter aspect, Das-Smaal & De Swart (1981) showed in a set of experiments that the degree of representativeness of features to their "central" value (e.g. the similarity of the shape of a flower leaf to the most typical form for that flower) influenced category learning, but that this effect was restricted to features that were relevant to the categorization decision. Relevancy was determined by frequency of occurrence in one category relative to occurrence in other categories. The discrepancy in effects of relevant versus irrelevant features was thought to be explained best by an interaction of selective attention, guided by expectations, and the knowledge that is built up (Das-Smaal, 1986; Das-Smaal *et al.*, 1987). In the course of learning, the information that has been gathered is used to direct attention to features deemed relevant. Features considered irrelevant to category abstraction may be represented incompletely or may not be remembered at all. Without abstraction of relevant features this phenomenon is hard to explain. In fact, it is not clear what constitutes a category at all when only exemplars are represented, as Smith and Medin (1981) pointed out. Also, there is a problem about what is done when only summary information is given and no exemplars are experienced. Categories may be learned by being stated rather than via examples. Some kind of abstraction seems to be required, although it may certainly be the case that some individual exemplars are also remembered and used in subsequent categorical judgments (see for instance Brooks, 1987).

Perhaps an exemplar view combined with a new approach to memory processes can overcome the objections mentioned above. Recently, Hintzman (1986) put forward such a theoretical approach. Although he does not preclude the existence of abstract memory representations, Hintzman maintains that they have no special status or function. It is assumed that each specific event to which one attends, is encoded as a new episodic memory trace. The theory differs from other approaches to memory in that information is abstracted from concrete experiences at the time of retrieval rather than during learning. A new experience gives rise to a process of activating or reminding of other similar experiences, and it is the summed contribution of these activated traces that is used to categorize the new experience, by analogy. The point is that abstraction does occur, but only momentarily, as a result of activation by a new experience. It need not be built up during learning, and stored explicitly in memory.

A variant of this view was put forth by Das-Smaal (1986, pp. 121-123) to account for the effects of context and of the flexible use of categories found in her experiments. A distinction was suggested between the information that is stored to represent a category, and procedures for using that information. Categories do not exist in memory in a ready-made format. It was contended that in order to be able to account for a number of phenomena regarding categorization behavior, the constituent parts of the category information should be separately available in memory. During learning, it appears that attention, guided by our schemas or theories and expectancies, does play an important role in determining which aspects will be represented and which ones will not. Next, computational mechanisms make use of the information in a flexible way, as needed, dependent on task requirements or the utilization goal. For instance, a certain feature can be highlighted in a metaphor or primed by advance activation of a category to which it is related. Whenever a category member is experienced, it gives rise to a momentary activation pattern, which in turn results, in some specified way, in the activation of categories. Once a category is learned, excitatory and inhibitory relations among correlated features and contrasting categories, respectively, are assumed to contribute to the resulting activation pattern. Furthermore, due to contextual factors certain features may be temporarily more important than others. That is, these features receive some extra activation, while others may become simultaneously inhibited. Contextual factors are used in a broad sense here, and may also imply the "internal" context of features activated by thought or experience. Regarding context, it was Barsalou (1982) who made the distinction between context-independent and context-dependent information in categories. Context-independence arises automatically from a high frequency of occurrence of features in a category as experienced by a person. Context-independent information is activated on all occasions. Context-dependent information, on the other hand, refers to properties that are activated only in certain contexts. In his more recent work, Barsalou (1985, 1987) also stresses that representations should not be viewed as invariant structures, but as dynamic ones that vary across contexts. Whenever a category is used, it is constructed in working memory from knowledge in long-term memory. Context-independent information may automatically be part of it. What other information is used depends on the situation. We shall return to the biasing influence of context later in this chapter.

2.3 Contrasts and combinations

The theoretical views described above were only summarily dealt with here. Many variants have been proposed. Prototype models may differ in what is assumed to be abstracted. This is an important issue regarding biases. What exactly do we learn from experience with a category's members and nonmembers? One could assume, for instance, that feature experience frequencies are abstracted (e.g. Neumann, 1974), or distances among exemplars (Posner & Keele, 1970), or an integrated pattern of propositions plus an allowable set of transformations (Bransford & Franks, 1971). Among feature frequency models a further distinction is possible between representation of simply independent features and their frequencies (Kellogg, 1981), and representation of both simple and conjoint frequencies of features (Hayes-Roth & Hayes-Roth, 1977). In addition, combinations of the views have been proposed, for example, in Medin & Schaffer's context-theory (1978) in which both exemplar-representation and a more abstract representation are hypothesized. Frequency- and prototype-distance models appear not to be mutually exclusive (Das-Smaal & De Swart 1984, 1986). A continuous variable dimensional value like the color red may show up within a category (e.g. apples) in several different variants. These variants of the value red differ in the degree to which they are typical of the category. However, there are also differences regarding the relative frequency with which the value red occurs within the category. The frequency variable has its own clear and specific effects on category learning and its results, independent of the effect of variant typicality. In short, frequency- and prototype-distance models that account for effects of frequency and variant typicality, respectively, may be complementary to one another. A frequency model that assumes continuous dimensions divided into intervals, and as with discrete values, representations of frequencies of these intervals, accounts for both effects mentioned (Das-Smaal & De Swart, 1986). Moreover, such a model unites the classical and the prototype approaches in the sense that the same principles of categorizing underlie both "classical" categories, with defining features, and categories with features that occur often but not necessarily always among category members. "Definingness" in a strict sense should then be replaced by some related but less rigid measure of variability, such as the degree to which a feature is representative of a category. This avoids the problem of whether or not people are able to specify defining features or whether or not they ultimately have defining features represented. At the same time it has the advantage of accounting for some cases of differences among category members as to how good an example they are. The problem that then remains is just what it is that determines variation in representativeness or typicality. Differences in typicality have been found in a wide range of category types, and it appears to be a very important variable in relation to all kinds of tasks having to do with categories. It is therefore important to understand the constraints related to these phenomena. We shall turn to this subject now. In the next sections we take up some basic issues in preferential ways of learning and using categories.

3. BIASES IN CATEGORIZATION

3.1 Prototypes

From the preceding discussion it can be concluded that the work of Rosch and her associates has suggested very interesting issues that are relevant to preferential biases in categorization.

Although it should be kept in mind that many other factors are involved in learning and determining category membership, one very important factor has been described in the literature as the usage of prototypes. It has been argued that many categories are represented by prototypes, with exemplars varying in their similarity to the prototype. The implication is that people do not act equally toward all members of a category. A chair is a better example of furniture than a lamp or a piano. People are much faster at determining category membership for typical members than they are for atypical members. Also, when thinking about a category, typical members come to mind with a higher probability than less typical ones. Prototypes may furthermore serve as reference points in similarity judgments (Rosch, 1975). We tend to select the prototype as a referent, and the other item as a subject. The reference point is the stimulus which other stimuli are seen "in relation to". For instance, we say that a panther resembles a cat rather than that a cat resembles a panther. Tversky and Kahneman (1983) showed how the use of such reference points can lead to biases in reasoning. People employ typical cases as the basis for their judgments, and inferences are made from typical to non-typical members. Also the results of comparisons between items appear not to be invariant with respect to the directionality of the comparison. Thus, the choice of a reference point may influence one's judgment. The issue of biases in reasoning is discussed extensively elsewhere in this book.

What sorts of biases are further implied by prototype theory? Following her observation that natural categories are organized around a prototype, two general and basic principles for the formation of categories were given by Rosch (1978). The first is a functional one and asserts that people are biased to develop category systems that provide maximum information with the least effort. The second regards the structure of the information thus provided, and implies that the perceived world comes to us as structured information rather than as arbitrary, or unpredictable features.

The first principle states that categorizing is aimed at what Rosch calls "cognitive economy". This means on the one hand that categories should preserve information about the environment as much as possible, but on the other hand -- to minimize cognitive load -- that they should reduce the infinite differences among stimuli to manageable proportions. In other words, the aim is to maximize the information accounted for, and to keep at a minimum the number of categories that have to be distinguished. Economy also implies a maximization of intercategorical differences, so that the categories will be as distinctive as possible.

In this context, the conception of hierarchical organization of categories is of relevance. Rosch claims that natural categories are hierarchically related. Larger categories usually contain a number of smaller ones. Three levels are distinguished. Superordinate categories, such as clothing, contain basic-level categories (e.g. trousers), which in turn contain subordinate categories (e.g. Levi's). Cognitive economy is maximized at the basic level. The basic level is the level at which the objects share the most features with the other members of the category. At lower levels, the features of members of a particular category most frequently overlap with features of members of other categories. That is, most of them are distinctive, whereas at superordinate levels there is less total similarity within a given category because there are relatively few features common to all members. Basic levels categories like car, table, or bird are the ones that languages tend to denote by single nouns, the ones that children learn first, and the ones we use most often in talking about the objects around us.

Rosch maintains that her claims concerning a basic level of abstraction can be formalized in terms of cue validity (e.g. Beach, 1964). Cue validity is a probabilistic concept indicating the predictive validity of a feature of a category, and is based on the frequency of occurrence of features in both the focal category and other categories. A category exemplar with a high total cue validity is differentiated more from other categories than one of lower total cue validity. Cue validity is maximized at the basic level.

Besides the first principle of categorization, cognitive economy, another principle was mentioned by Rosch. This second principle asserts that the environment is perceived as possessing a correlational structure. The combinations of what we perceive as features are not equiprobable. Rather, some features co-occur more than others. Rosch argues that we tend to form categories that mirror the structure perceived in the environment, although she recognizes in her later work that it may also be true that this structure is something that is imposed on regularities in nature by our conceptualizing minds (Rosch, 1978). Evidence that people indeed are sensitive to cue validity and correlated features will be presented separately in later sections.

In the description of Rosch's basic principles of categorization, it was assumed that category representations can be decomposed into features. However, it should be mentioned that in the literature there is uncertainty as to exactly what features are. Features can be concrete or abstract, and they may be categories themselves too. Their perception may be the result of a built-in constraint of our perceptual system, or they may count as a feature due to theory-related constraints. That is, our theories determine them as relevant because of convenience to our way of thinking about the world (Medin & Wattenmaker, 1987). We will not go further into the question of what these features are. This, however, is an important issue that requires further clarification.

Application of the prototype approach has not been limited to object categories. Research has extended it to different domains, where it serves as a useful organizational framework. Psychological constructs like intelligence (Neisser, 1979) and emotion (Fehr & Russell, 1984) show a graded structure, as do categories of personality traits (Buss & Craik, 1983; Smid & Zwinderman, 1986) and categories of social perception (Cantor, Mischel, & Schwartz, 1982; Cohen, 1983). In applying prototypes to personality research, various traits, features, behaviors, or situations are distinguished according to the degree of typicality relative to a concept from this domain. It appears, for instance, that people consider excitement as a more typical feature of emotion than boredom. In this way, researchers try to put order in the contents of these constructs, which are often vague. In addition, the results are used to design questionnaires (e.g. Broughton, 1984; De Jong, 1988; Visser & Das-Smaal, in preparation). Here the prototype idea has also been shown to be a fruitful approach. The biases mentioned in the beginning of this paragraph concerning the non-equivalence of different category members are put into practice in these domains.

To summarize, two important categorization principles mentioned by Rosch serve to bias or constrain the categories that we form. These are striving for cognitive economy, which is maximized at the basic level, and sensitivity to correlated features. The implication of the principles of categorization is that to increase distinctiveness among categories, categories tend to be formed around prototypes. Prototypes represent the features that are not only the most representative of the category members, but also the least representative of nonmembers. People do not consider all category members as equivalent, but are biased to focus on prototypes in a number of cognitive activities.

3.2 Schemas and levels of cognitive functioning

A kind of bias stemming from prototypes that has not been mentioned is the development of specific expectations resulting from the activation of a prototype. In this respect, the term schema is often used. The use of a schema framework has become widespread in psychological research. At the same time interest is now growing in complex, composite categories (Millward, 1980; Mandler, 1984; Medin & Smith, 1984). Research on categorization has been dominated by the use of simple categories. However, these categories can be further integrated and related to each other to form more complex organizations of knowledge. Such elaborated, interconnected knowledge structures are most generally referred to as schemas, although the schema notion implies also an active organizing principle. Not just categories can be coded by schemas. Events, stories and scenes can also be represented by schematic forms of organization. Prototypes can be considered as a sort of schema. To a certain extent some variation in the objects that might fit a particular schema is allowed, and constraints as to what the typical features are also exist. Schemas in general can be conceived of as conceptually organized clusters of knowledge. They lead to expectancies that guide both the comprehension of what we perceive, and the planning and execution of our actions. People can instantiate schemas in the absence of external evidence, and these instantiations are referred to as their default values (see for instance Mandler, 1984). In making category judgments, people often rely on prototypical schemas.

Social stereotypes can be considered to function in an analogous way. Social stereotypes are frequently used to stand for a category as a whole, for instance people of a certain race. As such, they may lead one to jump to conclusions about individual members of that category. This may be viewed as the application of features to the entire category that actually apply only to the stereotype. In terms of schema activation it can be stated that only partial evidence can activate a whole stereotype schema. This also implies the activation of features that are not part of the current evidence, by default.

Likewise, schemas in general are used to guide the interpretation of what we perceive. It is a commonplace observation that people tend to perceive what they expect to perceive. In this way, knowledge puts important constraints on perception and categorization. The idea that stimulus analysis is guided by expectations is central to schema theories (Neisser, 1967). Current experiences are assimilated with past experiences, and what we subsequently perceive is influenced by these experiences. Thus, perception contains the memory of earlier encounters, as Arnold (1984) put it. As a result, the accounts of stimuli may be different from the stimuli that evoked them. Das-Smaal's (1986) findings on the effects of typicality-range experience, which will be described in the next paragraph on category boundaries, serve to illustrate this point. In a category learning experiment, two groups of subjects differed in the range of typicality over which the features of the stimuli were varied during learning. The groups subsequently appeared to respond in different ways to the same inter-stimulus distance. The "psychological" distance between the stimuli from two categories was larger when a small rather than broad range of typical variants was experienced during learning. This shows that prior experience can make a significant difference in typicality judgments and in subjective intercategorical distance.

The observation that people use schemas to infer that certain unobserved and unmentioned features must be present is an obvious source of bias. It was Bartlett (1932) who applied the concept of schema to the domain of memory, in particular in relation to text comprehension. He pointed out that part of what we remember is our understanding of what we have experienced. Perception and memory should be viewed as constructive and reconstructive activities, respectively. Bartlett showed that subjects recalling a rather bizarre story exhibit systematic inaccuracies in memory. They distort the story to fit their own stereotype schemas. It is currently assumed that schemas are produced by an interaction between what we experience in the actual world, and our expectations driven by other activated schemas. Both factors can be very personal, especially regarding the affective components. Therefore, one might expect especially this source of bias to have a personal aspect. However, in the case of socially determined stereotypes, as no personal experience is implied and the stereotype schema is typically shared by a group of persons, this personal aspect by definition pertains to certain groups of people.

Recently, much attention has been paid to the importance of the relationship between category representation and other levels of cognitive functioning. For the higher levels, for instance, Murphy and Medin (1985) and Medin and Wattenmaker (1987) emphasize the influence of people's theories about the world on categorization. This idea has already been touched on earlier in this chapter. In their opinion, categories should be viewed as embedded in coherent, integrated knowledge systems, that in turn are based on perception, memory, and imagination. Categories are coherent insofar as they fit into someone's beliefs and knowledge. Implicit theories constrain our understanding of relationships both within and among categories to a high degree. That is not to say that these are the only constraints affecting categorization. Although theoretical factors are believed to be important, other factors, like the use of primarily perceptual information, play a role too. According to Murphy and Medin, feature models, in principle, can encompass the important aspects of general knowledge, but this has not been accomplished yet.

The schema notion has also been described in terms of activation patterns in a parallel distributed processing approach based on the brain metaphor (Norman, 1986). This approach has recently become increasingly popular. Low-level processing structures are assumed to result in higher order regularities in the following way. Information is represented by features, and each type of feature is represented by some unit in memory. Each stimulus attended to gives rise to a pattern of activation of units, and each time, this changes the strength of the connections among simultaneously occurring units. A unit may be involved in various events. Therefore, memory is said to be distributed across feature units. Units are functionally autonomous, but may show parallel activity. Categories are coded in terms of patterns of connections among units. A category has its own specific pattern of activation that mirrors the regularities of experience. Schemas or prototype patterns develop if many similar activation patterns are superimposed. This results in a composite trace that functions like an abstraction. Schemas are not fixed, permanent cognitive structures, but temporary patterns of activity. A similar pattern can be reinstated at another time. However, each re-creation may differ somewhat from the previous ones. The schema is constructed anew for each occasion by combining past experiences with biases and activation levels that result from the actual stimulus and context (Norman, 1986). Therefore, in this approach, category flexibility is accounted for, since the context is part of the constellation of activated units. This is an obvious advantage in view of the biasing influence of context. That is, categories may vary as the current context varies. For future theoretical development, it is an interesting idea that the assumption of distributed memory representation is not necessarily at odds with models of categorization (Knapp & Anderson, 1984; Oden, 1987).

3.3 Category width and context

The description of categories in terms of prototypes with some degree of allowable variation among members raises the problem of defining category width. How much discrepancy is allowable among members of a category? Can constraints on boundaries between categories arise as a result of learning alone, and if so, how? What are our biases in this respect? This issue has not received much attention. Some research has indicated that the category width variable experienced during learning does have an impact on subsequent categorization. It is also true that perceived boundaries are context-dependent. Furthermore, as we shall see, category ranges themselves can have a biasing influence on similarity judgments among items.

Although categories are formed in bounded representations, boundaries are not always clear. It was Rosch (e.g. 1975) who stressed this aspect of natural categories. On the other hand, in the prototype approach, the question of the ultimate range of discrepancies among category exemplars that a person is willing to accept was left open. Differences in degree of membership among exemplars was the main point dealt with in this approach. However, it is obvious that people may differ as to the ranges of their concepts of few, several, many, or a lot, for example. What are the minima, and what the maxima of each of these concepts? Context may be of relevance here. People do not always have to react to an object or an event simply in ways typical of some category to which the object or event belongs. People also have to respond to new aspects of the total situation at hand. They must be able to adapt their behavior to specific circumstances. Context may provide a clue as to how the category should be apprehended in a particular situation. Context may cause certain features to be temporarily considered as more important than they would otherwise be.

A study by Labov (1973) provides an illustration of fuzzy boundaries and contextual influence. Labov studied the boundary between the cup and bowl categories. He was interested in which items subjects would call cups, and which ones they would call bowls, when given a series of items. The items increased in their ratio of width to depth. Subjects were asked to classify pictures of these objects. With increasing relative width, there was a gradual shift in responses from cup to bowl, but there was no clear-cut point where subjects stopped using cup. Even more interesting was that when subjects were asked to imagine the object placed on a table and filled with food, more bowl responses were given, although the same gradual shift appeared from cup to bowl. Thus, boundaries not only can be unclear, but they can also be influenced by the context in which something is placed. This shows that boundaries are flexible. Regarding individual variation in boundaries, McCloskey and Glucksberg (1978) showed that boundaries for categories like fruit or disease vary among people. It even appeared that people themselves are not stable in their opinion. Many subjects changed their mind about the categorization of some atypical members when tested one month later.

Although there seems to be some instability about category width both among and within people, a number of researchers have shown in a variety of tasks that, broadly, people can be qualified as either relatively narrow, or relatively wide categorizers. That is, some people consistently categorize atypical items in the same category, whereas others consistently categorize these same items in different categories (Detweiler, 1978). According to Detweiler, these individual differences are the result of developmental, cultural, and experiential factors. Detweiler also suggests some effects of category width as an individual difference variable of behavior. For instance, narrow categorizers show better face recall, adjust less well to different cultures, are more ethnocentric, and make different attributions to foreigners than to non-foreigners (Detweiler, 1975, 1978).

Regarding biases, it is important to note that differences in category width may yield different categorizations of a particular object or event. Thus, the same stimulus may be interpreted differently by different persons. The foregoing remarks on context and individual differences concerned boundaries once they are formed. Das-Smaal and De Swart (1984, 1986) addressed the question of boundary formation. Their aim was not to ascertain individual differences due to styles of categorization. Rather, the possible influence of experience on category width was investigated. The results not only showed how category ranges can be determined by learning experience, but also how range experience influences the perceived within-category variation in typicality. In the learning phase, exemplars of a focal and a contrast category were composed of either a small range of typical variants only, or a broad range of both typical and atypical exemplars. As predicted, in contrast to narrow range experience, experience with a broad range of variants resulted in a relatively large extension of the focal category. It also resulted in better categorization of novel atypical focal category exemplars. This confirms Detweiler's idea that experience may determine the subsequent way of categorizing. Another effect shown in the studies by Das-Smaal and De Swart concerned the judged distance between categories and subjective typicality differences within categories. It appeared that small as compared with broad range experience had the effect of polarizing the focal and the contrast category. Thus, the same physical distance was judged differently, depending on range of experience. This polarization effect counts as a source of bias in category learning, caused by the category width experienced during learning. In addition, a recurrent finding in the experiments was the compression of the focal category following broad range experience. That is, members of the same category were felt to look more alike in this condition. To test whether this effect was indeed due to the width of the category, and not to the greater number of variants, in the later study a third range condition was added. This condition implied broad range experience, but with a limited number of differing variants. The results were essentially the same as with a broad range, indicating that the range was the crucial factor.

To summarize, the influence of the category range experience factor can be characterized in the following way. Learning experience with a broad rather that narrow range of variants results in enhanced discriminability across a category boundary, and in decreased discrimination within a category. A similar range bias has been reported lately by Conner, Land and Booth (1987).

3.4 Feature frequency

In paragraph 3.1 on prototypes, it was stated that Rosch suggested maximization of cue validity to be a determining factor regarding the internal structure of a category. According to this principle, the best examples of categories are those exemplars that have the most in common with other members of the same category, and share the least with contrasting categories. It is the latter addition that constitutes an important research issue. When people learn a category, do they use information on contrasting categories at all when forming the focal category? And if so, then what information do they use, and what effect does it have on the category to be learned?

It is often assumed that in learning a category, the learner keeps track of the frequencies with which features occur within that category. Indeed, there is substantial evidence for accumulation of frequency information during category learning (e.g. Neumann, 1974; Goldman & Homa, 1977; Hayes-Roth & Hayes-Roth, 1977; Chumbley *et al.*, 1978; Kellogg *et al.*, 1978). One could argue, then, that in a category representation the features represented are the ones that occur most often among category members. This is in fact what Neumann (1974) for instance proposed in his attribute-frequency model. According to his model, a category prototype is formed that contains the maximal number of the most frequently experienced features.

Occurrence of features within one category certainly is of relevance in the representation of that category. However, it could also be the case that precisely the same things that distinguish one category from other categories make up the most important information. In this view, when a category is being learned, the learner pays particular attention to those features that provide the sharpest contrast with other related categories. These so-called distinctive features may then become central to the category representation. In that view, a category representation not only contains information on occurrence in the focal category, but also on occurrence in contrasting categories. The latter information can of course only be learned when contrasting categories are experienced in addition to the category to be learned. In the latter view, contrasting categories are useful in a learning phase because they provide information on the occurrence of features outside the focal category. Distinctiveness from other categories is not always accounted for in categorization models. It is however accounted for in models that have cue validity as a critical term, such as the property-set model proposed by Hayes-Roth and Hayes-Roth (1977). As mentioned earlier, cue validity can be defined as the frequency with which a cue, or a feature, occurs in one category, divided by the total frequency of that cue across all categories. Cue validity thus takes into account resemblance within a category, as well as distinctiveness from contrasting categories.

In the category learning tasks of several experiments by Das-Smaal and De Swart (1984, 1986), feature frequency was varied in both the focal and contrasting categories. As for occurrence in the focal category, a high frequency of features was assumed to facilitate categorization performance. In addition, the influence of occurrence in a contrasting category was investigated extensively on the hypothesis that the frequency of the features in the contrasting category would affect categorization performance and representativeness ratings. Category exemplars were expected to be categorized more easily and judged as more representative as the frequency of their composing features in the contrasting category was lower during learning. This effect was studied both by varying frequency in the contrasting category and by either including or not including a contrast category in the learning task.

Frequency of occurrence of features in the focal category, as compared to their occurrence in the contrasting category, was expressed in terms of cue validity. The question was what do subjects learn from this state of affairs about the representativeness of a feature; i.e. to what degree is it used and relied upon in subjective judgments following learning?

The influence of the frequency of feature occurrence among contrasting category items on categorizing focal category exemplars was a very robust finding, coming from converging sources. It was demonstrated repeatedly, employing a variety of learning paradigms, testing methods, and dependent variables. In all experiments, high total cue validity of exemplars facilitated learning and classification following learning. It also appeared that when learning had to take place from focal category exemplars only, without experiencing a contrast category, categorization was guided by the focal frequency of the features.

When features with the same cue validity but with different frequencies of occurrence in the focal category were compared, focal frequency showed its own influence, apart from cue validity. Features of higher frequency were judged as more typical. However, distinctive features -- that is, features that occur frequently and exclusively in the focal category -- appeared to be the most typical and most important ones to classification following learning. Delaying testing proved that this was all the more true with the passing of time.

The apparent conclusion from the findings is that people collect considerable knowledge of the frequency with which features occur in particular categories during learning. This knowledge is not restricted to just the focal category, but also extends to the contrast category. People are able to use this knowledge in order to evaluate relative frequencies, that is, occurrence in one category as compared with occurrence in a contrast category.

3.5 Correlation among features

Another controversial issue regarding frequency of occurrence is whether or not subjects learn about each feature in an independent manner from the other features. The question is: Does frequency of co-occurrence of features affect category formation? Furthermore, if people are sensitive to correlations among features, does this mean that all correlations are registered?

Rosch *et al.* (1978) have claimed that the perceived world comes as structured information rather than as arbitrary, equiprobable co-occurring features. Features that characterize the members of natural categories are often correlated, and these conjoint frequencies are mirrored in the categories that people form. The suggestion that conjoint frequency affects categorization was also made by Hayes-Roth and Hayes-Roth (1977) and by Medin and Schaffer (1978). Murphy and Medin (1985) also state that correlations are represented in mental categories, although they need not always be based on empirical relations; they can also mirror a person's ideas about the relationships.

The point of correlational structure has also been emphasized by Anderson (1985). Anderson describes the structure of natural categories as a schematic structure, and asserts that schemas represent (among other things) our knowledge about how dimensional values tend to go together to define objects. Thus according to Anderson, it is the interrelational structure, the configuration of dimensional values, rather than just a list of values that defines a category.

On the other hand, Kellogg (1981) argues that stimulus features are independently represented, and that only simple frequencies are counted. Kellogg's study, a category learning experiment, failed to produce conjoint frequency effects. In his study, however, the features upon which conjoint frequency was varied were all irrelevant to categorization. The possibility remains that conjoint frequency effects will show up when relevant features are concerned. This question was tackled by Das-Smaal & De Swart (1986). In that study, it appeared indeed that the effectiveness of conjoint frequency shows up when at least one of the features is important to categorization. Conjoint frequency increases the judged typicality of an irrelevant feature when it correlates with a feature that is important to categorization. This explains why Kellogg (1981) did not find any conjoint frequency effect using features that were all irrelevant to categorization. Moreover, this is another indication of a source of bias in category learning.

The assumption that effects of conjoint frequency are constrained to situations in which at least one of the joint features is relevant to or characteristic of the category also solves a major problem noted both by Reed (1982) and by Medin and Smith (1984) and Medin and Wattenmaker (1987). They argue that it is not probable that all possible correlations among features are taken into account, because they are too numerous. The condition of relevancy of at least one of the conjoint features may in this respect appear to be a very important constraint on the correlations that are taken into account.

Stated in terms of the view put forward by Das-Smaal (1986) mentioned at the end of paragraph 2.2 in this chapter, the following may apply. Excitatory relations between two irrelevant features do nothing significant to increase the activation of the category. However, when at least one relevant feature is concerned, the increased activation of the category is the result, facilitating categorization and increasing typicality judgments. The irrelevant feature activates the category by way of its relationship to the highly characteristic feature.

Another interesting potential explanation bears on the Gestalt claim that "the whole is more than the sum of the parts". It may be speculated that the formation of interrelations by conjoint frequency, and the resulting additional activation through mutual excitation, gives rise to some of these effects. The results of the present study indicate that correlation of features enhances the typicality of the "whole" stimulus containing the correlated features, relative to non-redundant combination. This shows that the conjoint occurrence of values during learning can cause the whole to be perceived as more than the sum of its parts. The existence of a predominant manner of organizing the parts of a stimulus may, in a comparable way, be due to special interrelations between features in memory, whether they are learned or physiologically "pre-wired". The effects of learning the interrelations have been established in the study by Das-Smaal and De Swart (1986).

3.6 Selective attention

From the study conducted by Das-Smaal and De Swart (1986) mentioned above, it appeared that typicality of an irrelevant feature increases only if this feature correlates with a feature that is important to categorization. This could explain why Kellogg (1981) found no effect of correlations, because in his experiment the features between which the correlations were varied were both irrelevant to the categorization decision. Results from other experiments by Das-Smaal and De Swart (1981) indicate that with a well-defined category like the one Kellogg used, attention is directed at relevant features, not at irrelevant ones, or if so to a lesser degree. Typicality effects were restricted to features that were relevant to the categorization decision. It should be noted here that typicality and relevance in these experiments were not the same thing. Typicality, as it was conceived of in these experiments, had to do with the range of a feature, whereas relevance had to do with its probability of occurrence. Features may vary in their frequency of occurrence, and this affects categorization as discussed previously. However, continuous features may also vary in the typicality of the variants with which they occur. For instance, the feature red may occur with a certain range of variants of this color, and these variants are not equally typical of red. It appeared that category learning was facilitated by high rather than low typical feature variants when relevant features were concerned. No such difference was found with irrelevant features. The discrepancy may be explained by the existence of an interaction between the selectivity of attention and the knowledge that is built up (e.g. Das-Smaal *et al.*, 1987). In the beginning of category learning, there is no category information available in memory to direct attention to specific features. The analysis of the stimulus is data driven, probably influenced by the salience of features. However, in the course of learning, the analysis becomes more "top down", guided by hypotheses or expectations. The information that has been gathered is used to direct attention to features deemed relevant, the latter being guided, e.g. by frequency. Finally, the learner will end up with detailed representations of information on at least the focused features. The information that is stored includes various details about frequencies as well as information on their typicality ranges, as has been shown in studies by Das-Smaal and De Swart (1984, 1986). Features considered irrelevant to category abstraction may be represented incompletely or may not be remembered at all. This is what a recognition experiment by Nickerson and Adams (1979) indicated. These authors showed that whereas knowledge of color and size of a U.S. penny is good, knowledge of its visual details is poor. Features such as color and size are relevant to the distinction between a penny and other coins. Additional details, on the other hand, are irrelevant to the purposes for which pennies are employed. Hence they apparently do not receive much attention.

3.7 Hypothesis testing

In the foregoing, selective attention was mentioned as a source of bias. Expectations or hypotheses contribute to directing attention to what is deemed relevant. Since this places a constraint on the kind of information that is gathered regarding a category, it specifies the bias of selective attention.

The idea of hypothesis testing when learning categories was studied in experiments following the work by Bruner described earlier. The investigations dealt with the kind of hypotheses subjects entertained during learning and how these hypotheses were revised on the basis of feedback. A number of models have been developed regarding the way in which a subject might choose among a pool of hypotheses. Early models (Restle, 1962; Bower & Trabasso, 1964) assumed that the subject begins a learning task with a universe of hypotheses from which he draws one. This hypothesis dictates his response. The subject retains the hypothesis if his response turns out to be correct, but discards it and selects another one if his response is wrong. A refuted hypothesis is supposed to be returned to the pool. Later experiments refuted the assumption of no memory for previously tested hypotheses (e.g. Levine, 1966).

A more elaborate version of hypothesis-testing theory was formulated by Levine (1969). This theory assumes that the subject begins a learning task by sampling a subject from the universe of hypotheses. He or she then takes one hypothesis from the subset as the working hypothesis upon which his or her response is based. The working hypothesis is retained if the response is correct. If the response is wrong, the working hypothesis is discarded, and a new working hypothesis is chosen from the subset. If the subset is empty, the subject takes a new subset of several hypotheses and chooses a new working hypothesis from this subset. Other hypotheses from the subset are updated simultaneously after each feedback. These hypotheses are eliminated from the subset when they turn out to be wrong. In contrast to the no-memory assumption of the early models, Levine's theory assumed that subjects do eliminate hypotheses from the pool, at least for some time. In the course of time, subjects may forget that some of the hypotheses were not confirmed. These hypotheses then become part of the pool again. The weight of empirical evidence supports Levine's theory (e.g. Levine, 1975; De Swart & Das-Smaal, 1976, 1979a, 1979b; Bourne, Dominowski, & Loftus, 1979; Das-Smaal & De Swart, 1981; De Swart, Kok, & Das-Smaal, 1981).

De Swart and Das-Smaal (1976, 1979a, 1979b) argued that Sokolov's (1969) model of the orienting reflex is a neural analogue of the cognitive process of hypothesis testing. The authors (1979b) found that the amplitude of the Skin Conductance Response (SCR) was higher when feedback during category learning confirmed a categorization response in which subjects had low confidence as compared with one on which they reported high confidence. Refutation revealed the opposite result: SCR was higher following a high-confidence categorization than following a low-confidence categorization. A study by De Swart, Kok and Das-Smaal (1981) replicated these results with the amplitude of P300 as a measure of changes in the probability of hypotheses.

The findings of Das-Smaal (1986) mentioned earlier on the storage and use of information on frequency of occurrence in a categorization learning task can be reconciled with hypothesis testing theory by the assumption that frequency information is used to select hypotheses. Haygood *et al.* (1970) e.g., and more recently Kellogg (1980), have provided evidence of this view. One special theory on hypothesis formation, Levine's (1969) multi-hypothesis sampling theory, was supported in a study by Das-Smaal and De Swart (1981). An illustration of multi-hypothesis sampling in a real-life situation was given by Reed (1982). Reed describes how hypothesis testing is applied to a task that involves diagnosing medical problems. Results on how physicians attempt to diagnose a disease agree with Levine's theory. It appeared that physicians start to form hypotheses early in the examination. Then they monitor a subset of about three hypotheses at a time, and this subset remains fairly constant through different stages of examination. The results show that more than one hypothesis or expectation can be active and evaluated simultaneously.

3.8 Task demands

Ease of interpretation of new impressions is not the only function of categories. Categories are essential in thinking and problem solving. They also make conversation more easy, at least if the involved categories have names and are subject to general agreement within a culture. Regarding categories, many cognitive judgments are possible. Das-Smaal and De Swart (1984, 1986) presented their subjects, following category learning, with test tasks of different kinds. These were e.g. a categorization task and a task in which subjects were asked to choose the more representative one of two stimuli for the focal category. Differential performance on focal category exemplars, in particular those of lower total cue validity, was found consistently between these two tasks in all experiments. The results were taken to suggest that frequency in a contrast category is weighted more heavily in categorization tasks than in tasks that require judgments of the representativeness of exemplars relative to the focal category. Unlike in categorization decisions, differences among categories are not so much the issue in the latter kind of task. Therefore, distinctiveness, as determined by occurrence in contrast categories, is deemed less important and receives less weight. This is not to say that the effect of occurrence in a contrast category is not important at all in a paired comparisons task. Studies have shown that it is important to a certain extent. However, the fact that the above disparity showed up repeatedly points to the relevance of task analysis. Comparable effects of task demands in categorization versus subsequent justification of a categorization decision are described by Medin and Smith (1984) and Landau (1984). Also, the results of Kemler Nelson (1984) and of Ward and Scott (1987) are very interesting in this respect. They found the same task effect as described above. In a traditional category learning task (intentional learning), subjects appeared especially to attend to the feature that maximized the difference between categories. For certain reasons, the intentional learning task in this study was compared with an incidental learning task. The latter task in fact boiled down to pairwise comparisons. The instruction was to determine which one of two exemplars was most typical of a category. In this task, in contrast to the categorization task, subjects granted no extra weight to the distinctive feature.

A post-hoc interpretation of the findings on a more elementary level of theorizing may serve to integrate the results of Das-Smaal (1986) on frequency of occurrence in different tasks. Suppose the following. As a part of a momentary activation pattern, features activate categories in which they have occurred to the extent corresponding to the previously registered frequency of occurrence. Thus, a common feature occurring often in the focal category but also in a contrast category activates both the focal and the contrast category, although the focal category will be activated more. A distinctive feature, however, occurring often in the focal but never in a contrast category, activates only the focal category. It is furthermore assumed that contrasting categories inhibit each other when they are activated, a suggestion made by Wickelgren (1981) and Rumelhart and McClelland (1982). It follows that the net activation of the focal category will be higher due to a distinctive feature than to a common feature, even if they occur with the same frequency in the focal category, as in the Das-Smaal (1986) studies. In a categorization task, the decision may then involve determination of which one of the categories at stake is activated most. A pairwise comparison task, on the other hand, concentrates on the focal category. It involves determining which one of the two exemplars is more strongly related to the focal category. In such a task the focal category may be primed, and this diminishes the relative influence of inhibition by way of the contrast category.

The issue of task-dependence is, of course, not a new one in psychology. The point is that not much attention has been paid to the influence of task differences on the conclusions reached in much of the experimental research on categorization. One interesting implication of the findings on task demands mentioned above is that information on frequency in the focal category and information on frequency in contrast categories has to be separately accessible in memory, and not already joined in the measure of cue validity. Cue validity, then, may be either computed when necessary, or stored in addition to frequencies in the focal and contrast categories. Perhaps what is required is a distinction like the one made by Miller and Johnson-Laird (1976) between categories and procedures for using categories. Having knowledge of occurrence in contrasting categories does not imply that this knowledge is always used to perform.

Thus the contention that a category consists of certain features is not enough. It does not say how those features are weighted or integrated when people use that category in some particular situation. An account is needed of the manner in which a whole is constructed from the parts in each case. And as for storage, the information must be stored in such a way that it enables us to use that information in a flexible way. Task and contextual factors may give rise to a kind of momentary relevant activation pattern in which certain values may temporarily be of more importance than others.

The results indicate the importance of knowing which demands a task places on the subject. The effect of task difference shows that category information can be used in a flexible way. It is not very likely that all aspects of category representation can be investigated in one single type of task. Which demands are made by the various tasks in the current research paradigms is not yet clear. Systematic analysis and categorization of tasks is therefore necessary. In the future, this may prevent research partiality.

4. CONCLUSIONS

In this chapter, a number of important theoretical approaches have been presented regarding the question of how people are biased in the way they divide up the world into categories. In addition, experimentally established sources of bias were analysed. Overall, perhaps the most striking observation in this overview is the fact that many of the constraints on human categorization behavior described in the principal theoretical alternatives seem to be implicated. The point is that they do not all occur under the same conditions, and there is a lack of knowledge about the way in which they are combined. There is evidence indicating the abstraction of schemas or prototypes, learning of feature frequencies both of the focal and contrast categories from the same domain, sensitivity to correlations among features, exemplar representation, category width effects, hypothesis testing, effects of selective attention, and for the very important role served by context, task and background knowledge and ideas. These processes serve to create a knowledge base regarding categories that can be activated in a flexible way. Categorization behavior then may be regarded as the result of some computational process which appears to be influenced by the cognitive constructs that we personally entertain in relation to the to-be-categorized object, the task at hand or the goal of the categorization activity, and the context in which it takes place. All of the factors mentioned above can bias our categorization behavior in certain directions. It is a matter of further investigation to determine their boundary conditions.

5. REFERENCES

Anderson, B.F., Deane, D.H., Hammond, K.R., McClelland, G.H., & Shanteau, J.C. (1981). Concepts in judgement and decision research. New York: Praeger Publishers.

Anderson, J.R. (1985). Cognitive psychology and its implications. New York: W.H. Freeman and Company.

Armstrong, S.L., Gleitman, L.R., & Gleitman, H. (1983). What some concepts might not be. Cognition, 13, 263-308.
Arnold, M.B. (1984). Memory and the Brain. Hillsdale, NJ: Lawrence Erlbaum Associates.
Barsalou, L.W. (1982). Concext-independent and context-dependent information in concepts. Memory and Cognition, 10, 82-93.
Barsalou, L.W. (1983). Ad hoc categories. Memory and Cognition, 11, 211-217.
Barsalou, L.W. (1985). Ideals, central tendency and frequency of instantiation as determinants of graded structure in categories. Journal of Experimental Psychology: Learning, Memory and Cognition, 11, 629-654.
Barsalou, L.W. (1987). The instability of graded structure. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization. Cambridge: Cambridge University Press.
Bartlett, F.C. (1932). Remembering. London and New York: Cambridge University Press. Beach, L.R. (1964). Cue probabilism and inference behavior. Psychological Monographs, 78, No. 582.
Bourne, L.E., & O'Banion, K. (1969). Memory for individual events in concept identification. Psychonomic Science, 16, 101-103.
Bourne, L.E., Dominowski, R.L., & Loftus, E.F. (1979). Cognitive processes. Englewood Cliffs, NJ: Prentice Hall.
Bourne, L.E. (1982). Typicality effects in logically defined categories. Memory and Cognition, 10, 3-9.
Bower, G., & Trabasso, T. (1964). Concept identification. In R.C. Atkinson (Ed.), Studies in mathematical psychology. Stanford, CA: Stanford University Press.
Bransford, J.D., & Franks, J.J. (1971). Abstraction of linguistic ideas. Cognitive Psychology, 2, 331-350.
Brooks, L.R. (1987). Decentralized control of categorization: the role of prior processing episodes. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization. Cambridge: Cambridge University Press.
Broughton, R. (1984). A prototype strategy for the construction of personality scales. Journal of Personality and Social Psychology, 47, 1334-1346.
Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). A study of thinking. New York: Wiley.
Buss, D.M., & Craik, K.H. (1983). The act frequency approach to personality. Psychological Review, 90, 102-126.
Cantor, N., Mischel, W., & Schwartz, J.C. (14). A prototype analysis of psychological situations. Cognitive Psychology, 45-77.
Chumbley, J.L., Sala, L.S., & Bourne, L.E. (1978). Bases of acceptability ratings in quasi-naturalistic concept-tasks. Memory and Cognition, 6, 217-226.
Cohen, C.E. (1983). Inferring the characteristics of other people: categories and attribute accessibility. Journal of Personality and Social Psychology, 44, 34-44.
Conner, M.T., Land, D.G., & Booth, D.A. (1987). Effect of stimulus range on judgements of sweetness intensity in a lime drink. British Journal of Psychology, 78, 357-364.
Das-Smaal, E.A., & De Swart, J.H. (1981). Typicality, physiological activity and concept identification. Acta Psychologica, 47, 1-24.
Das-Smaal, E.A., & De Swart, J.H. (1984). Variation within categories. Acta Psychologica, 57, 165-192.
Das-Smaal, E.A., & De Swart, J.H. (1986). Effects of contrasting category, conjoint frequency and typicality on categorization. Acta Psychologica, 62, 15-40.
Das-Smaal, E.A. (1986). Variation within categories. Amsterdam: Free University Press.

Das-Smaal, E.A., De Leeuw, L., & Orlebeke, J.F. (1987). Is er iets mis met de aandacht van het schoolkind. Pedagogische Studiën, 64, 1-15.
De Jong, P.F. (1988). An application of the prototype scale construction strategy to the assessment of student motivation. Journal of Personality, 56, 487-508.
De Swart, J.H., & Das-Smaal, E.A. (1976). Relationship between SCR, heart rate and information processing. Biological Psychology, 4, 41-49.
De Swart, J.H., & Das-Smaal, E.A. (1979a). Physiological activity in an attribute learning task. Acta Psychologica, 43, 57-69.
De Swart, J.H., & Das-Smaal, E.A. (1979b). Orienting reflex and uncertainty reduction in a concept learning task. In J.F. Orlebeke (Ed.), The orienting reflex in humans. Hillsdale, NJ: Lawrence Erlbaum Associates.
De Swart, J.H., Kok, A., & Das-Smaal, E.A. (1981). P300 and uncertainty reduction in a concept identification task. Psychophysiology, 18, 619-629.
Detweiler, R.A. (1975). On inferring the intentions of a person from another culture. Journal of Personality, 43, 591-611.
Detweiler, R.A. (1978). Culture, category width, and attributions; A model-building approach to the reasons for cultural effects. Journal of Cross-Cultural Psychology, 9, 3, 259-284.
Dominowski, R.L, & Wethenick, N.E. (1976). Inference processes in conceptual rule learning. Journal of Experimental Psychology: Human Learning and Memory, 2, 1-10.
Fehr, B., & Russell, J.A. (1984). Concept of emotion viewed from a prototype perspective. Journal of Experimental Psychology: General, 113, 464-486.
Goldman, D., & Homan, D. (1977). Integrative and metric properties of abstracted information as a function of category discriminality, instance variability, and experience. Journal of Experimental Psychology: Human Learning and Memory, 3, 375-385.
Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behavior, 16, 321-338.
Haygood, R.C., Harbert, T.L., & Omlor, J.A. (1970). Intradimensional variability and concept identification. Journal of Experimental Psychology, 83, 216-219.
Hintzman, D.L. (1986). "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93, 411-428.
Hull, C.L. (1920). Quantitative aspects of the evolution of concepts. Psychological Monographs, 123.
Hunt, E.B., & Hovland, C.I. (1960). Order of consideration of different types of concepts. Journal of Experimental Psychology, 59, 220-225.
Kellogg, R.T., Bourne, L.E., & Ekstrand, B.R. (1978). Feature frequency and the acquisition of natural concepts. American Journal of Psychology, 91, 212-222.
Kellogg, R.T. (1980). Feature frequency and hypotheses testing in the acquisition of rule-governed concepts. Memory and Cognition, 8, 297-303.
Kellogg, R.T. (1982). Feature frequency in concept learning: What is counted? Memory and Cognition, 9, 157-163.
Kemler Nelson, D.G. (1984). The effect of intention of what concepts are acquired. Journal of Verbal Learning and Verbal Behavior, 23, 734-759.
Knapp, A.G., & Anderson, J.A. (1984). Theory of categorization based on distributed memory storage. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 616-637.
Labov, W. (1973). The boundaries of words and their meanings. In R.W. Shuy (Ed.), New ways of analyzing variation in English. Washington D.C.: Georgetown University Press.
Lakoff, G. (1986). Women, fire, and dangerous things: What categories tell us about the nature of thought. Chicago: University of Chicago Press.

Landau, B. (1982). Will the real grandmother please stand up. Journal of Psycholinguistic Research, 11, 47-62.
Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. Journal of Experimental Psychology, 71, 331-338.
Levine, M. (1969). Latency-choice discrepancy in concept learning. Journal of Experimental Psychology, 82, 1-3.
Levine, M. (1975). Hypothesis Testing: a cognitive theory of learning. Hillsdale, NJ: Lawrence Erlbaum Associates.
Mandler, G. (1984). Mind and body: psychology of emotion and stress. New York: W.W. Norton & Company.
Mandler, J.M. (1984). Stories, scripts and scenes: aspects of schema-theory. Hillsdale, NJ: Lawrence Erlbaum Associates.
Martin, R.C., & Caramazza, A. (1980). Classification in well-defined and ill-defined categories: Evidence for common processing strategies. Journal of Experimental Psychology: General, 109, 320-353.
McCauley, R.H. (1987). The role of theories in a theory of concepts. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization. Cambridge: Cambridge University Press.
McCloskey, M.E., & Glucksberg, S. (1978). Natural categories. Well-defined or fuzzy sets?. Memory and Cognition, 6, 462-472.
Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. Psychological Review, 85, 207-238.
Medin, D.L., & Smith, E.E. (1984). Concepts and concept formation. Annual Review of Psychology, 35, 113-138.
Medin, D.L., & Wattenmaker, W.D. (1987). Category cohesiveness, theories and cognitive archeology. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization. Cambridge: Cambridge University Press.
Miller, G.A., Galanter, E.H., & Pribram, K.H. (1960). Plans and the structure of behavior. New York: Holt, Rinehart & Winston.
Miller, G.A., & Johnson-Laird, P.N. (1976). Language and perception. Cambridge, MA: Harvard University Press.
Millward, R.B. (1980). Models of concept formation. In W.E. Montagne (Ed.), Aptitude, learning and instruction, Vol. 2. Hillsdale, NJ: Lawrence Erlbaum Associates.
Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. Psychological Review, 92, 289-316.
Neisser, U. (1967). Cognitive psychology. New York: Appleton-Century-Crofts.
Neisser, U. (1979). The concept of intelligence. Intelligence, 3, 217-227.
Neisser, U. (1987). Introduction: the ecological and intellectual bases of categorization. In U. Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual factors in categorization. Cambridge: Cambridge University Press.
Neumann, P.G. (1974). An attribute frequency model for the abstraction of prototypes. Memory and Cognition, 2, 241-248.
Nickerson, R.S., & Adams, M.J. (1979). Long-term memory for a common object. Cognitive Psychology, 11, 287-307.
Norman, D.A. (1986). Reflections on cognition and parallel distributed processing. In The PDP Research Group (Ed.), Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge: MIT Press.
Oden, G.C. (1987). Concept, knowledge and thought. Annual Review of Psychology, 38, 203-227.
Posner, M.I., & Keele, S.W. (1970). Retention of abstract ideas. Journal of Experimental Psychology, 83, 304-308.
Reed, S.K. (1982). Cognition. Theory and applications. Monterey: Brooks/Cole Publishing Company.

Restle, F. (1962). The selection of strategies in cue learning. Psychological Review, 69, 329-343.

Reznick, J.S., & Richman, C.L. (1976). Effects of class complexity, class frequency and pre-experimental bias on rule learning. Journal of Experimental Psychology: Human Learning and Memory, 2, 774- 782.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press.

Rosch, E. (1975). Cognitive representation of semantic categories. Journal of Experimental Psychology: General, 3, 192-233.

Rosch, E. (1975). Cognitive reference points. Cognitive Psychology, 7, 532-547.

Rosch, E. (1978). Principles of categorization. In B.B. Lloyd (Ed.), Cognition and categorization. Hillsdale, NJ: LEA.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braen, P. (1978). Basic objects in natural categories. Cognitive Psychology, 8, 382-439.

Rumelhart, D.E., & McClelland, J.L. (1982). An interactive model of context effects in letter perception: II. Psychological Review, 89, 60-94.

Smid, N.G., & Zwinderman, A.H. (1986). Representatie van een eigenschappen-bij-situaties taxonomie in prototypische gedragsbeschrijvingen. Nederlands Tijdschrift voor de Psychologie, 41, 91-103.

Smith, E.E., & Medin, D.L. (1981). Categories and concepts. Cambridge: Harvard University Press.

Sokolov, E.N. (1969). The modeling properties of the nervous system. In I. Maltzman (Ed.), A handbook of contemporary Soviet psychology. New York: Basic Books.

Tversky, A., & Kahneman, D. (1983). Probability, representativeness, and the conjunction fallacy. Psychological Review, 90, 293-315.

Vandierendonck, A. (1988). Typicality gradient in well-defined artificial categories. Acta Psychologica, 69, 61-81.

Visser, M.R.M., & Das-Smaal, E.A. (in preparation). Assessment of impulsiveness using the prototype scale construction strategy.

Ward, T.B., & Scott, J. (1987). Analytic and holistic modes of learning family-resemblance concepts. Memory and Cognition, 15, 42-54.

Wickelgren, W.A. (1981). Human learning and memory. Annual Review of Psychology, 32, 21-52.